

KOREAN PATENT ABSTRACTS

(11)Publication number: **1020040042231 A**

(43)Date of publication of application:
20.05.2004

(21)Application number: **1020020070418**

(71)Applicant: **ELECTRONICS AND
TELECOMMUNICATIONS
RESEARCH INSTITUTE**

(22)Date of filing: **13.11.2002**

(72)Inventor: **CHOI, SEUNG GWON**

(54) DEVICE AND METHOD FOR TRANSLATING DO VERB OF AUTOMATIC KOREAN-CHINESE TRANSLATING SYSTEM BASED ON VERB PHRASE PATTERN

(57) Abstract:

PURPOSE: A device and a method for translating a do verb of an automatic Korean-Chinese translating system based on a verb phrase pattern are provided to improve a performance of the translation system while reducing a number of the verb phrase patterns needed from the translation system by analyzing a structure of a Korean sentence and converting/generating the Chinese. CONSTITUTION: A morpheme analyzer(101) receives/analyzes a morpheme of the Korean sentence. A sentence structure analyzer(102) analyzes the sentence structure by using the verb phrase pattern. A translated sentence converter(103) converts/generates a translated Chinese expression by using an analysis result of the sentence structure. A morpheme generator(104) generates the final Chinese sentence by receiving the result from the translated sentence converter and generating a Chinese vocabulary morpheme.

(19)대한민국특허청(KR)
(12) 공개특허공보(A)(51) 。 Int. Cl.⁷
G06F 17/28(11) 공개번호 10-2004-0042231
(43) 공개일자 2004년05월20일(21) 출원번호 10-2002-0070418
(22) 출원일자 2002년11월13일(71) 출원인 한국전자통신연구원
대전 유성구 가정동 161번지(72) 발명자 서영애
대전광역시유성구지족동열매마을대우아파트303동1003호김영길
대전광역시중구오류동삼성아파트1동1005호양성일
대전광역시유성구신성동대림두레아파트110동1008호김창현
서울특별시용산구후암동448-156류철
대전광역시서구관저동구봉마을아파트512동1001호홍문표
대전광역시유성구도룡동383-2과기원교수아파트2동102호노윤형
대전광역시유성구신성동208-9301호이기영
대전광역시서구만년동1-1초원아파트106동411호최승권
대전광역시유성구어은동99한빛아파트113동1101호박상규
대전광역시유성구송강동한마을아파트105동1305호(74) 대리인 권태복
이화익

심사청구 : 있음

(54) 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의번역장치 및 방법

요약

본 발명은 동사구 패턴에 기반한 한중 자동 번역 시스템에서 '하다' 동사의 번역 장치 및 방법에 관한 것으로, 본 발명의 목적은 'X를 하다' 구문의 처리를 위한 별도의 '하다' 동사의 동사구 패턴을 구축하지 않게 하여 시스템에서 요구되는 동사구 패턴의 수를 줄이면서도 고품질의 번역이 가능하게 하는 것이다.

본 발명에 따른 한국어 구조 분석기는 입력 결과에서 'X를 하다' 구문이 존재하는 지를 인식하고, 이를 'X하다' 구문으로 바꾸는 제 1단계; 데이터베이스에서 'X하다' 동사구 패턴을 가져와서 제1단계의 입력으로 들어온 'X를 하다' 구문의 제약조건을 만족하는 최적의 'X하다' 동사구 패턴을 선택하는 제 2단계; 'X하다' 동사구 패턴을 이용하여 한국어 입력문의 구문구조를 분석하는 제 3단계; 제 3단계에서 구문 분석된 'X하다' 구문의 결과를 'X를 하다' 구문 구조로 치환하는 제 4단계를 수행한다. 또한, 본 발명에서의 대역문 변환기는 이를 해결하기 위해 'X'가 관형절의 수식을 받는 'X를 하다' 구문이 입력되었는지를 인식하는 제 5단계; 수식을 받는 경우 'X를 하다' 구문이 어떤 유형에 속하는지를 판단하는 제 6단계; 'X를 하다'의 유형에 따라 관형어/절의 처리를 수행하는 제 7단계를 수행한다.

대표도

도 2

색인어

한중 자동번역, 동사구 패턴, '하다' 동사, 구조 분석기, 대역문 변환기

명세서

도면의 간단한 설명

도 1은 본 발명이 이용되는 동사구 패턴에 기반한 한중 자동 번역 시스템을 보인 블록구성도,

도 2는 본 발명에 따라 한국어 구문 분석기가 'X를 하다' 구문에 대해 구조 분석을 수행하는 과정을 나타내는 순서도,

도 3은 본 발명에 따라 대역문 변환기가 유형1의 'X를 하다'에 대한 대역문 변환을 수행하는 과정을 나타내는 순서도,

도 4는 본 발명에 따라 대역문 변환기가 유형2의 'X를 하다'에 대한 대역문 변환을 수행하는 과정을 나타내는 순서도,

도 5는 본 발명에 따라 대역문 변환기가 유형3의 'X를 하다'에 대한 대역문 변환을 수행하는 과정을 나타내는 순서도.

* 도면의 주요 부분에 대한 부호의 설명 *

100;한중 자동 번역 시스템 101;형태소 분석기

102;구문 분석기 103;대역문 변환기

104;형태소 생성기

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 한국어를 중국어로 자동으로 번역해주는 컴퓨터를 이용한 자동번역기술(MT)에 관한 것으로, 더욱 상세하게는 동사구 패턴에 기반한 한중 자동번역 시스템에서 'X하다'와 'X를 하다' 구문간의 상관성을 이용하여 'X를 하다' 구문을 'X하다' 동사에 대한 동사구 패턴을 이용하여 번역을 수행하는 장치 및 방법에 관한 것이다.

최근들어, 컴퓨터를 이용한 자동 번역 시스템분야에서 예제 기반 자동 번역이나 패턴 기반 자동 번역과 같이 데이터에 기반한 번역방법론에 대한 연구가 많이 이루어지고 있다. 규칙 기반 방식(Rule-based MT)과 더불어 자동번역 시스템 구축의 주요한 방법론인 데이터 기반 방식(Data-driven MT)은 규칙 기반 방식과는 달리 대량의 데이터를 필요

로 하여, 구축된 데이터의 양과 질이 시스템의 성능을 결정한다.

동사구 패턴에 기반한 한중 자동 번역 시스템은 데이터에 기반한 자동 번역 시스템으로서, 동사구 패턴을 기반으로 하여 한중 자동 번역을 수행하므로 동사구 패턴의 양과 질이 번역 시스템의 성능에 결정적인 영향을 미친다. 여기서, '동사구 패턴'은 한국어 용언의 쓰임을 기술한 하위 범주화 틀(subcategorization frame)의 확장된 형태로서, 한국어 구조 분석 정보와 중국어 대역 정보가 함께 기술되어 있다.

동사구 패턴과 구조 분석만을 위해 기술되는 일반적인 하위 범주화 정보와의 가장 큰 차이점은 필수 논항뿐만 아니라 대역 부분에 영향을 미치는 수의 논항(optional argument) 및 부사 등과 같은 부가어들도 기술된다는 점이다.

다음 표 1은 '가다' 용언에 대한 동사구 패턴 정보의 예이다.

[표 1]

가다1: A=사람!가 B=건축물!에 가!다 > A 去: v B[그는 법원에 갔다]
가다7: A=시간!가 가!다 > A 법去: v 了[하루가 가다/세월이 가다]

상기 표 1은 '가다' 용언에 대한 동사구 패턴 정보의 예를 기술한 그림이다. '>'의 왼쪽부분은 한국어 동사구 패턴으로 한국어에 대한 정보이고, 오른쪽 부분이 이에 대응하는 중국어 정보를 기술한 중국어 동사구 패턴이다.

한국어 동사구 패턴부에서 공백으로 구분된 각 항은 용언의 격슬롯 및 부가어 정보들이며, '[' 안의 내용은 해당 동사구 패턴에 대한 예문이다. 동사구 패턴의 예에서 A, B 등은 용언의 격슬롯을 의미하며, 격슬롯에 대한 의미제약이 의미코드 '사람', '건축물', '시간' 등에 의해 기술되어 있다.

A, B 등의 기호 없이 표현하여 어휘제약을 기술 할 수도 있다. 번역하고자 하는 한국어 용언구가 한국어 동사구 패턴의 제약조건을 모두 만족할 경우, 오른쪽의 중국어 동사구 패턴 정보를 이용하여 중국어로 변환, 생성을 수행한다.

일반적으로 격들의 구축은 어려운 작업으로 알려져 있다. 한중 동사구 패턴은 격들의 구축비용에 중국어 대역문의 구축비용까지 더해짐으로써 매우 고비용, 고난이도의 작업이다. 따라서 의미적으로 같은 동사구일 경우, 동사구 패턴 정보를 공유함으로써 패턴 구축비용을 줄일 필요가 있다.

종래의 기술에서는 'X하다'와 'X를 하다' 구문은 서로 다른 동사구 패턴을 이용하여 번역을 수행하였다. 즉, 'X하다'의 경우는 'X하다' 동사의 동사구 패턴을, 'X를 하다'의 경우는 '하다'동사의 동사구 패턴을 이용하여 번역을 수행하였다.

예를 들어, 다음 예문 (1)과 (2)의 문장에 대해서 각각 다음 표 2와 표 3의 동사구 패턴을 이용하여 번역을 수행하였다.

예문(1): 그는 마침내 결정했다.

예문(2): 그는 마침내 결정을 했다.

그러나 예문 (1)과 (2)의 문장에서 보듯이 '결정하다'가 '결정을 하다'로 바뀌었을 뿐, 이 두 구문은 의미적으로 동일하며 취하는 격 성분 정보도 동일하다. 다음 표 2 및 표 3에서 보면, 표 2와 표 3의 동사구 패턴도 '결정하!다'와 '결정!을 하!다' 부분을 제외하고는 동일하다. 표 2는 '결정하다' 구문을 위한 '결정하다' 동사구 패턴의 예이고, 표 3은 '결정을 하다' 구문을 위한 '하다' 동사구 패턴의 예이다.

[표 2]

결정하다12: A=사람!가 마음!를 결정하!다 > A 決心: v [그가 마음을 결정하다]

결정하다13 : A=사람!가 B=방법!를 결정하!다 > A 決定:v B [그가 방침을 결정하다]

[표 3]

하다35 : A=사람!가 마음!를 결정!을 하!다 > A 決心:v [그가 마음을 결정을 하다]

하다36 : A=사람!가 B=방법!를 결정!을 하!다 > A 決定:v B [그가 방침을 결정하다]

따라서, 하나의 동사구 패턴을 공유하도록 하여 패턴의 중복 기술을 피하도록 하는 것이 바람직하다.

이와 관련된 종래 유사 기술로는 한 문장어구의 논항의 수와 형태를 결정짓는 중심어 개념을 도입한 패턴 기반 번역 방식에 있어서 서술성 명사를 목적으로 취하는 '하다' 동사 구문에 대해 서술성 명사를 중심으로 패턴 분석하는 방법이 제안된 바 있으나, 종래 기술에서는 서술성 명사와 결합한 기능동사 구문으로 그 범위를 한정하여 비서술성 명사에 대해서는 처리하지 못했으며, 구문 분석의 관점에서만 논하고 있어 목적 언어로의 변환 및 생성시의 문제점 및 이에 대한 해결책을 제시하지 못한다는 단점이 있었다.

발명이 이루고자 하는 기술적 과제

본 발명은 상기 문제점을 해결하고 상기 요구에 부응하기 위하여 제안된 것으로, 동사구 패턴에 기반한 한중 자동 번역 시스템에서 'X를 하다'와 'X하다' 구문의 상관성을 이용하여 'X를 하다' 구문이 'X하다' 동사구 패턴을 이용하여 한국어 구문 분석 및 중국어 변환 생성을 하도록 함으로써 시스템이 필요로 하는 동사구 패턴의 수를 줄이면서도 번역 시스템의 성능을 높이는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역장치 및 방법을 제공하는데 그 목적이 있다.

상기 목적을 달성하기 위하여 본 발명의 동사 번역장치는, 한국어 문장을 입력받아 이를 형태소 분석하는 형태소 분석기; 상기 형태소 분석결과를 받아 이를 구문 분석하되, 'X하다' 동사구 패턴을 이용하여 'X를 하다' 구문의 구조를 분석하는 구문 분석기; 상기 구문 분석결과를 이용하여 한국어 동사구들의 중국어 대역표현을 생성하되, 'X하다' 동사구 패턴을 이용하여 'X를 하다' 구문의 중국어 대역표현을 변환 및 생성하는 대역문 변환기; 및 상기 대역문 변환기의 결과를 입력받아 중국어 어휘 형태소를 생성하여 최종 중국어 문장을 생성하는 형태소 생성기;를 포함한 것을 특징으로 한다.

또한 상기 목적을 달성하기 위하여 본 발명의 동사 번역방법은, 한국어 문장을 입력받아 이를 형태소 분석하는 형태소 분석단계; 상기 형태소 분석결과를 입력받아 'X를 하다' 구문이 존재하는지를 인식하고, 이를 'X하다' 구문으로 바꾸는 제 1단계, 데이터베이스에서 'X하다' 동사구 패턴을 가져와서 상기 제 1단계의 입력으로 들어온 'X를 하다' 구문의 제약조건을 만족하는 최적의 'X하다' 동사구 패턴을 선택하는 제 2단계, 'X하다' 동사구 패턴을 이용하여 한국어 입력문의 구문구조를 분석하는 제 3단계, 상기 제 3단계에서 구문 분석된 'X하다' 구문의 결과를 'X를 하다' 구문 구조로 치환하는 제 4단계로 이루어진 구문 분석 단계; 상기 구문 분석결과를 이용하여 한국어 동사구들의 중국어 대역표현을 생성하는 대역문 변환단계; 및 상기 대역문 변환단계의 결과를 입력받아 중국어 어휘 형태소를 생성하여 최종 중국어 문장을 생성하는 형태소 생성단계;를 포함한 것을 특징으로 한다.

발명의 구성 및 작용

이하, 본 발명의 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다. 도 1은 본 발명에 적용되는 동사구 패턴을 이용한 한중 자동 번역 시스템을 개략적으로 나타낸 도면이다.

도 1을 참조하면, 본 발명의 동사구 패턴을 이용한 한중 자동 번역 시스템(100)은 한국어 문장을 입력받아 이를 형태소 분석하는 형태소 분석기(101), 형태소 분석결과를 받아 이를 구문 분석하는 구문 분석기(102), 구문 분석결과를 이용하여 한국어 동사구들의 중국어 대역표현을 생성하는 대역문 변환기(103) 및 대역문 변환기(103)의 결과를 입력

받아 중국어 어휘 형태소를 생성하여 최종 중국어 문장을 생성하는 형태소 생성기(104)를 포함한다.

이와 같이 구성된 한중 자동 번역 시스템에서 한국어 구문 분석기(102)가 본 발명에 의해 'X를 하다' 구문을 번역하는 방법은 도 2와 같다.

도 2는 한국어 구문 분석기(102)에서 본 발명에 따라 'X하다' 동사구 패턴의 한국어 동사구 패턴을 이용하여 한국어 구조 분석을 수행하는 순서도이다.

도면을 참조하면, 본 발명은 먼저, 형태소 분석 결과에 'X를 하다' 구문이 있는지 조사하여(S201), 있을 경우 입력된 'X를 하다' 구문을 'X하다'구문으로 치환하고(S202), 치환된 'X하다' 동사구의 분석을 위해 'X하다' 동사구 패턴들을 데이터베이스에서 가져와(S203), 그 중에서 입력문의 정보와 'X하다' 동사구 패턴의 한국어 동사구 패턴 조건을 비교하여 가장 적절한 하나의 'X하다' 동사구 패턴을 선택한다(S204).

선택된 'X하다' 동사구 패턴의 한국어 부분을 이용하여 구문 분석을 수행한 후(S205), 구문분석이 끝나면 치환된 'X하다'를 원래의 'X를 하다' 구문으로 복원하여 최종적인 구문 분석 구조를 출력한다(S206).

'X를 하다'에서 명사 'X'가 관형어나 관형절의 수식을 받지 않는 경우에는 중국어 변환 및 생성 과정은 일반적인 동사구 패턴에 기반한 변환 및 생성 절차와 동일하다.

그러나 'X를 하다'에서 명사 'X'가 관형어나 관형절의 수식을 받는 경우 'X하다' 동사구 패턴의 중국어 대역 동사구 패턴 정보를 이용하여 'X를 하다' 구문을 중국어로 변환, 생성 시에 문제가 발생하는데 예를 들면 다음과 같다.

예문 (3) : 그는 마침내 어려운 결정을 했다.

예문 (4) : 他終於作了不容易的決定。

상기 예문 (3)은 '결정하다'와 '어렵다'에 해당하는 두 개의 동사구 패턴을 이용하여 번역을 수행할 경우, 예문 (4)와 같은 중국어 대역문이 생성되지 않는다. 예문 (4)의 대역문으로 올바르게 번역되기 위해서는 '결정'에 해당하는 중국어 대역어, '決定'이 대역문에 표현되어 두 중국어 용언구의 결합이 가능하여야 한다.

그러나 표 2의 동사구 패턴에는 한국어 '결정'에 대응하는 중국어 '決定'이 중국어 동사구 패턴에 나타나지 않으므로, '어려운'에 해당하는 대역어가 결정을 수식하는 예문 (4)와 같은 전체 중국어 번역문을 생성할 수 없다.

본 발명에서는 이를 해결하기 위해서 한국어의 'X를 하다' 용언구에 대응하는 중국어의 번역을 다음 표 4와 같이 분류하고, 각 타입에 적합한 변환 및 생성 방법으로 번역을 수행한다.

본 발명에서 대역문 변환기(103)는 'X'가 관형어나 관형절의 수식을 받는 'X를 하다' 구문이 입력되었는지를 인식하는 제 1단계, 수식을 받는 경우 'X를 하다' 구문이 어떤 유형에 속하는지를 판단하는 제 2단계 및 판단된 'X를 하다'의 유형에 따라 관형어나 관형절의 처리를 수행하는 제 3단계를 수행한다.

'X를 하다' 구문의 유형 분류는 우선, 중국어 번역 결과에 따라 크게 중국어 기능동사와 결합하는 유형과 결합하지 않는 유형으로 분류하고, 기능동사와 결합 불가능한 동사들은 그 내부구조에 근거하여 다시 2가지 유형으로 하위분류하여 유형화하여 총 3가지로 분류한다. 다음 표 4는 'X를 하다'에 대한 중국어 번역 유형을 정리한 것이다.

[표 4]

유형		중국어 대역표현	한국어 'X를 하다'의 예
1	A	加以, 給以, 予以, 給予 와 결합 가능	우대를 하다
	B	作, 進行와 결합 가능	연설을 하다
	C	A,B 기능동사와 모두 결합 가능	비평을 하다
2		동빈구조	강의를 하다
3		기타	공부를 하다

상기 표 4에서 유형1은 한국어 'X를 하다'에 대응하는 중국어 대역표현이 '중국어 기능동사 + 'X'의 중국어 대역어'의 형태로 번역 가능한 경우이다.

유형1에 속하는 'X를 하다' 구문의 관형어/관형절 수식 문제를 처리하기 위해서는 'X'의 형태소 사전 정보에 'X'의 중국어 대역어휘 정보와 'X'의 중국어 대역 어휘 정보에 대응하는 중국어 기능동사 정보를 기술한다.

중국어의 기능동사는 그 의미적 특징에 근거하여 A유형-'加以, 給以, 予以, 給予'과, B유형-'作, 進行'으로 나뉜다. 중국어 명사와 결합 가능한 기능동사의 선택은 각 어휘의 개별적 특성에 의해 결정된다.

A유형과 결합 가능한 중국어 동사들은 의미 특징상 모 객체에 영향(to inflict)주거나 수여(to confer)의 의미를 지니고 있는 반면에, B유형과 결합 가능한 중국어 동사들은 어떤 사건을 처리(to deal with) 혹은 거행(to hold)의 의미를 지닌다. 그리고 A, B유형과 모두 결합이 가능한 중국어 동사들은 상기의 두 가지 의미 특성을 두루 갖추고 있는 어휘들이다.

형태소 사전은 한국어 형태소 어휘, 형태소 코드, 의미코드(SEM), 빈도수, 영어 대역정보 (EROOT), 중국어 대역어(CROOT), 중국어 기능동사 정보(CSVERB), 중국어 수량사 정보(NC) 등을 담고 있다. 다음 표 5는 '결정'에 대한 형태소 사전 정보의 예이다.

[표 5]

```
'결정'
120012 479
{
[(SEM 선택)(EROOT decision)(CROOT 決定)(CSVERB 作, 加以)]
[(SEM 유생물상태)(EROOT crystallization)(CROOT 결정(結晶))]
}
```

이상의 정보를 이용하여, 유형1에 속하는 'X를 하다'의 'X'가 관형어/절의 수식을 받는 경우, 대역문 변환기(103)의 처리 흐름은 도 3과 같다.

도 3은 본 발명에 따라 대역문 변환기가 유형1의 'X를 하다'에 대한 대역문 변환을 수행하는 과정을 나타내는 순서도이다.

도면을 참조하면, 대역문 변환기(103)의 입력에 'X'가 관형어나 관형절의 수식을 받는 'X를 하다' 구문이 포함되어있는지를 조사하여(S301), 포함된 경우 'X를 하다'의 유형정보를 알아보기 위해 한국어 'X'의 형태소 사전 정보를 검사한다. 만약 한국어 'X'의 형태소 사전 정보에 기능동사정보(CSVERB)가 있는 경우, 유형1로 간주한다(S302).

유형1이 밝혀진 경우, 한국어 'X'에 대한 형태소 사전 정보에서 중국어 대역어 정보(CROOT)와 중국어 기능동사 정보(CSVERB)를 가져오고(S303), 다음으로 'X하다'의 중국어 동사구 패턴을 데이터베이스에서 가져온다(S304). 동사구 패턴이 한국어 동사구 패턴과 이에 대응하는 중국어 동사구 패턴의 쌍으로 이루어져 있기 때문에 한국어 구문 분석기에서 가장 적합한 한국어 동사구 패턴을 선택하게 되면 자동으로 중국어 동사구 패턴도 선택이 되므로, 대역문 변환기에서 별도로 적합한 중국어 동사구 패턴을 선택하지 않아도 된다. 중국어 동사구 패턴을 가져온 후, 중국어 동사구 패턴의 동사를 'X'의 형태소 사전에서 가져온 중국어 기능동사 정보(CSVERB)와 중국어 대역어 정보(CROOT)를 이용하여 중국어 '형식동사+ 서술성명사'의 형태로 치환한다(S305).

이어 치환된 'X하다' 중국어 동사구 패턴을 이용하여 'X를 하다' 구문에 대한 중국어 대역표현을 생성한 후(S306), 기존의 관형어/절 처리 루틴을 호출하여 관형어/절을 생성한다(S307). S306에 의해 생성된 'X를 하다' 구문에 대한 중국어 대역표현에는 한국어 'X'에 대응하는 중국어 표현이 번역되어 나타나므로, 'X를 하다' 구문의 중국어 대역표현에서 'X'의 중국어 표현을 찾아 'X'의 중국어 표현을 S307에서 생성된 중국어 관형어/절이 수식하도록 중국어 대역표현 결과를 결합하여 최종적인 번역을 완료한다(S308).

유형2는 'X를 하다'의 중국어 대역어가 '중국어 기능동사 + 'X'의 중국어 대역어'의 형태'로는 번역가능하지 않으나, 동빈(動賓) 구조로 이루어진 경우이다. 동빈 구조란 '講課'와 같이 하나의 중국어의 동사의 내부 구조가 다시 동사와

목적어의 결합으로 이루어진 경우이다. 'X하다' 동사에 대한 중국어 대역표현이 동빈구조인 경우, 중국어 동사의 내부 목적어가 'X'의 중국어 표현에 대응하므로, 중국어 동사의 번역결과에 'X'의 번역이 나타난다.

동빈 구조 정보의 기술은 'X하다' 동사의 중국어 동사구 패턴에 기술한다. 즉, 중국어 동사구 패턴의 중국어 동사에 대해, 중국어 동사 중 동빈 구조의 내부 동사에 해당하는 어휘 뒤에 ':v'를 기술하고, 내부 목적어에 해당하는 어휘 뒤에 ':n'을 기술하여 내부 구조를 표시한다.

다음 표 6은 '식사하다'의 한중 동사구 패턴의 예로서, 예문의 '식사하다' 동사의 동사구 패턴에서 '식사하다'에 대응하는 중국어 동사는 '吃飯'이다. '吃飯'은 동빈구조의 동사로서 내부 동사인 '吃'와 내부 목적어인 '飯'으로 이루어져 있는데, 이를 중국어 동사구 패턴부에 ':v'와 ':n' 기호를 덧붙여 중국어부의 ':v'는 동빈 구조에서의 중국어 동사(吃), ':n'은 목적어(飯)임을 표시한다.

[표 6]

식사하다1: A=사람!가 식사하!다 > A 吃:v飯:n [아버지께서 식사하시다.]

도 4는 본 발명에 따라 대역문 변환기가 유형2의 'X를 하다'에 대한 대역문 변환을 수행하는 과정을 나타내는 순서도이다.

도면을 참조하면, 유형2의 경우는, 유형1에서의 치환과정을 거치지 않고, 'X하다' 동사구 패턴의 중국어 대역부에서 ':n'의 링크정보로 기술된 'X'의 중국어 대역어를 수식하도록 관형어/절 처리 루틴을 호출한다. 즉, 대역문 변환기(103)의 입력에 'X'가 관형어나 관형절의 수식을 받는 'X를 하다' 구문이 포함되어 있고(S301), 'X'의 형태소 사전 정보에 기능동사 정보가 없는 경우(S302), 'X하다' 동사의 중국어 동사구 패턴을 데이터베이스에서 가져온다(S310). 가져온 'X하다' 중국어 동사구 패턴에 ':v'와 ':n' 표기가 있는지 확인하여 동빈 구조인지를 판단한다(S311). 동빈 구조인 경우, 'X하다' 중국어 동사구 패턴을 이용하여 'X를 하다' 구문의 중국어 대역 표현을 생성한다(S312).

그리고 나서 기존의 관형어/절 처리 루틴을 호출하여 관형어/절을 생성한다(S307). S312에 의해 생성된 'X를 하다' 구문에 대한 중국어 대역표현에는 한국어 'X'에 대응하는 중국어 표현이 번역되어 나타나므로, 'X를 하다' 구문의 중국어 대역표현에서 'X'의 중국어 표현을 찾아, 'X'의 중국어 표현을 S307에서 생성된 중국어 관형어/절이 수식하도록 중국어 대역표현 결과를 결합하여 최종적인 번역을 완료한다(S308).

유형 3은 유형1,2에 속하지 않는 그 외의 경우이다. 유형3의 경우를 살펴보면 한국어에서의 관형어 혹은 관형절이 중국어 문장에서는 술부를 수식하는 보어로 나타나는 것이 일반적이다. 따라서, 본 발명에서는 유형3의 'X를 하다' 구문에서 'X'를 수식하는 관형어나 관형절을 중국어의 부사어나 부사절로 생성한다. 유형3에 대한 처리 흐름도는 도 5와 같다.

도 5는 본 발명에 따라 대역문 변환기가 유형3의 'X를 하다'에 대한 대역문 변환을 수행하는 과정을 나타내는 순서도이다.

도면을 참조하면, 대역문 변환기(103)의 입력에 'X'가 관형어나 관형절의 수식을 받는 'X를 하다' 구문이 포함되어 있고(S301), 'X'의 형태소 사전 정보에 기능동사 정보가 없으며(S302), 데이터베이스에서 가져온 'X하다' 동사의 중국어 동사구 패턴에서 중국어 대역동사가 동빈구조가 아닌 경우(S311), 유형 3으로 간주하여 이에 대한 처리가 이루어진다.

먼저, 'X하다' 중국어 동사구 패턴을 이용하여 'X를 하다' 구문의 중국어 대역 표현을 생성한다(S313). 그리고 나서 'X'를 수식하는 관형어나 관형절을 부사어나 부사절로 간주하여 중국어 부사어나 부사절에 해당하는 대역 표현으로 생성한다(S314).

최종적으로, S314에서 생성된 중국어 부사어나 부사절의 대역표현이 S313에서 생성된 'X를 하다' 구문의 중국어 대역표현에서 'X하다'에 해당하는 중국어 동사를 수식하도록 결합하여 최종 중국어 문장을 생성한다(S315).

다음 예문 (5)에 대한 대역문 (6)이 바로 그 예이며, 번역된 중국어 문장 (6)을 한국어로 다시 번역하면 예문 (7)과 같

다. 여기서, 한국어의 관형절 '늦은'이 중국어의 부사어 '늦게'로 번역되었음을 알 수 있다.

예문(5): 그는 늦은 출발을 했다.

예문(6): 他出發得很晚。

예문(7): 그는 늦게 출발했다.

이상의 예에서는 'X를 하다'의 'X'가 '결정', '출발'과 같은 서술성 명사에 대해서만 예를 들었다. 그러나, 본 발명에서 제안하는 'X하다' 동사구 패턴의 공유를 통한 한국어 구조 분석 및 중국어 변환/생성 방법은 '나무하다', '나무를 하 다'와 같이 'X'가 서술성 명사가 아닌 경우에도 동일하게 적용된다.

발명의 효과

이상에서 설명한 바와 같이 본 발명은 동사구 패턴에 기반한 한중 자동번역 시스템에서 'X를 하다' 구문의 번역을 위해 추가의 '하다' 동사구 패턴의 구축 없이 'X하다' 동사의 동사구 패턴을 이용하여 번역하는 방법을 제안하여 'X하다' 동사구 패턴의 적용 범위를 넓힘으로써 한중 자동 번역 시스템에서 필요로 하는 동사구 패턴의 수를 줄이면서도 번역 시스템의 품질은 향상시킬 수 있다.

이상에서 설명한 것은 본 발명에 따른 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역장치 및 방법을 설명한 하나의 실시 예에 불과한 것으로써, 본 발명은 상기한 실시 예에 한정되지 않고, 이하의 특허 청구의 범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변경 실시가 가능한 범위까지 본 발명의 기술적 사상이 미친다고 할 것이다.

(57) 청구의 범위

청구항 1.

한중 자동 번역장치에 있어서,

한국어 문장을 입력받아 이를 형태소 분석하는 형태소 분석기;

상기 형태소 분석결과를 받아 이를 구문 분석하되, 'X하다' 동사구 패턴을 이용하여 'X를 하다' 구문의 구조를 분석하는 구문 분석기;

상기 구문 분석결과를 이용하여 한국어 동사구들의 중국어 대역표현을 생성하되, 'X하다' 동사구 패턴을 이용하여 'X를 하다' 구문의 중국어 대역표현을 변환 및 생성하는 대역문 변환기; 및

상기 대역문 변환기의 결과를 입력받아 중국어 어휘 형태소를 생성하여 최종 중국어 문장을 생성하는 형태소 생성기;를 포함한 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역장치.

청구항 2.

제 1항에 있어서, 상기 구문 분석기는,

상기 형태소 분석 결과에 'X를 하다' 구문이 있는 경우 입력된 'X를 하다' 구문을 'X하다'구문으로 치환하고, 치환된 'X하다' 동사구의 분석을 위해 'X하다' 동사구 패턴들을 데이터베이스에서 가져와, 그 중에서 입력문의 정보와 'X하다' 동사구 패턴의 한국어 동사구 패턴 조건을 비교하여 가장 적절한 하나의 'X하다' 동사구 패턴을 선택하여 구문 분석을 수행한 후, 구문분석이 끝나면 치환된 'X하다'를 원래의 'X를 하다' 구문으로 복원하여 최종적인 구문 분석 구조를 출력하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역장치.

청구항 3.

제 1항에 있어서, 상기 대역문 변환기는,

'X'가 관형어나 관형절의 수식을 받는 'X를 하다' 구문에 대한 대역 표현 생성을 위해 중국어 기능동사정보와 중국어 대역동사의 동빈구조 정보를 이용하여 변환을 수행하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역장치.

청구항 4.

한중 자동 번역방법에 있어서,

한국어 문장을 입력받아 이를 형태소 분석하는 형태소 분석단계;

상기 형태소 분석결과를 입력받아 'X를 하다' 구문이 존재하는지를 인식하고, 이를 'X하다' 구문으로 바꾸는 제 1단계와, 데이터베이스에서 'X하다' 동사구 패턴을 가져와서 상기 제 1 단계의 입력으로 들어온 'X를 하다' 구문의 제약조건을 만족하는 최적의 'X하다' 동사구 패턴을 선택하는 제 2단계와, 'X하다' 동사구 패턴을 이용하여 한국어 입력문의 구문구조를 분석하는 제 3단계와, 상기 제 3단계에서 구문 분석된 'X하다' 구문의 결과를 'X를 하다' 구문 구조로 치환하는 제 4단계로 이루어진 구문 분석단계;

상기 구문 분석결과를 이용하여 한국어 동사구들의 중국어 대역표현을 생성하는 대역문 변환단계; 및

상기 대역문 변환단계의 결과를 입력받아 중국어 어휘 형태소를 생성하여 최종 중국어 문장을 생성하는 형태소 생성단계;를 구비한 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역방법.

청구항 5.

제 4항에 있어서, 상기 대역문 변환단계는

상기 'X'가 관형어나 관형절의 수식을 받는 'X를 하다' 구문이 입력되었는지를 인식하는 제 5단계;

관형어나 관형절의 수식을 받는 경우, 'X를 하다' 구문이 어떤 유형에 속하는지를 판단하는 제 6단계; 및

상기 판단된 'X를 하다'의 유형에 따라 관형어나 관형절의 처리를 수행하는 제 7 단계;를 포함하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역방법.

청구항 6.

제 5항에 있어서, 상기 제 6단계는

'X를 하다' 구문의 중국어 번역 결과에 따라 크게 중국어 기능동사와 결합하는 제1 유형과, 중국어 기능동사와 결합하지 않고 동빈(動賓) 구조로 이루어진 제 2유형, 기타의 제 3유형으로 분류하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역방법.

청구항 7.

제 6항에 있어서, 상기 제 6단계에서 분류결과 제 1유형이면,

상기 제 7단계는

한국어 'X'에 대한 형태소 사전 정보에서 중국어 대역어 정보(CROOT)와 중국어 기능동사 정보(CSVERB)를 가져오는 단계;

'X하다'의 중국어 동사구 패턴을 데이터베이스에서 가져오는 단계;

중국어 동사구 패턴의 동사를 'X'의 형태소 사전에서 가져온 중국어 기능동사 정보(CSVERB)와 중국어 대역어 정보(CROOT)를 이용하여 중국어 '형식동사+ 서술성명사'의 형태로 치환하는 단계;

상기 치환된 'X하다' 중국어 동사구 패턴을 이용하여 'X를 하다' 구문에 대한 중국어 대역표현을 생성한 후, 기존의 관형어/절 처리 루틴을 호출하여 관형어/절을 생성하는 단계; 및

상기 생성된 'X를 하다' 구문의 중국어 대역표현에서 'X'의 중국어 표현을 찾아 'X'의 중국어 표현을 중국어 관형어/절이 수식하도록 중국어 대역표현 결과를 결합하는 단계;를 포함하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역방법.

청구항 8.

제 6항에 있어서, 상기 제 6단계에서 분류결과 제 2유형이면,

상기 제 7단계는

'X하다' 중국어 동사구 패턴을 이용하여 'X를 하다' 구문의 중국어 대역 표현을 생성하는 단계; 및

기존의 관형어/절 처리 루틴을 호출하여 관형어/절을 생성하고, 'X를 하다' 구문의 중국어 대역표현에서 'X'의 중국어 표현을 찾아, 'X'의 중국어 표현을 중국어 관형어/절이 수식하도록 중국어 대역표현 결과를 결합하는 단계;를 포함하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역방법.

청구항 9.

제 6항에 있어서, 상기 제 6단계에서 분류결과 제 3유형이면,

상기 제 7단계는

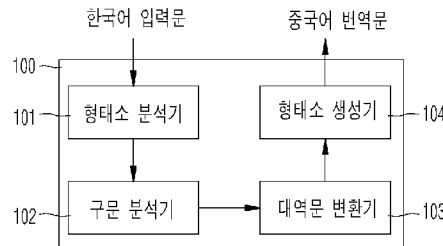
'X하다' 중국어 동사구 패턴을 이용하여 'X를 하다' 구문의 중국어 대역 표현을 생성하는 단계;

'X'를 수식하는 관형어나 관형절을 부사어나 부사절로 간주하여 중국어 부사어나 부사절에 해당하는 대역 표현으로 생성하는 단계; 및

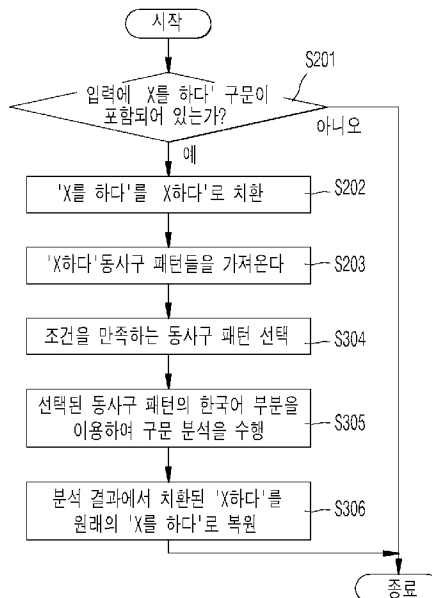
상기 중국어 부사어나 부사절의 대역표현이 'X를 하다' 구문의 중국어 대역표현에서 'X하다'에 해당하는 중국어 동사를 수식하도록 결합하는 단계;를 포함하는 것을 특징으로 하는 동사구 패턴 기반 한중 자동 번역 시스템의 '하다' 동사의 번역방법.

도면

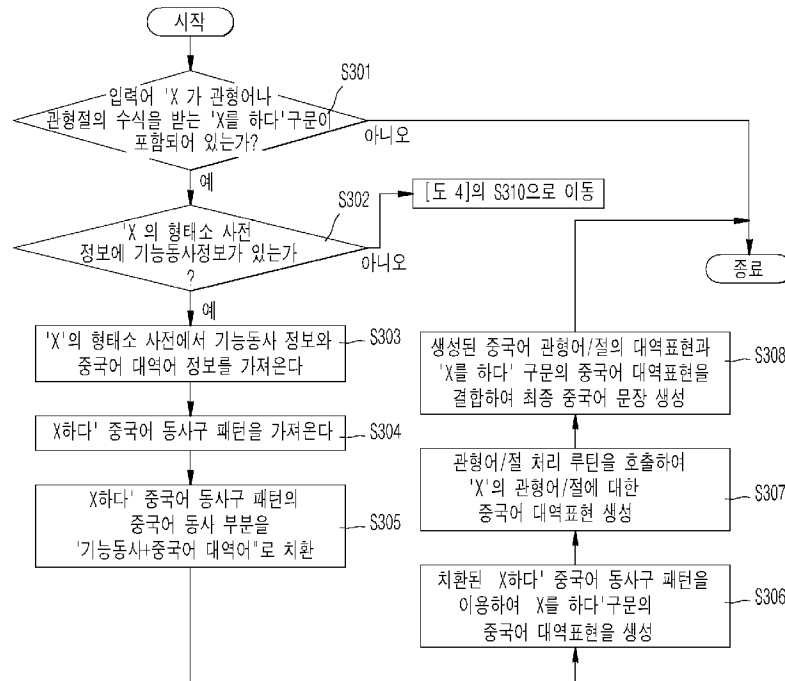
도면1



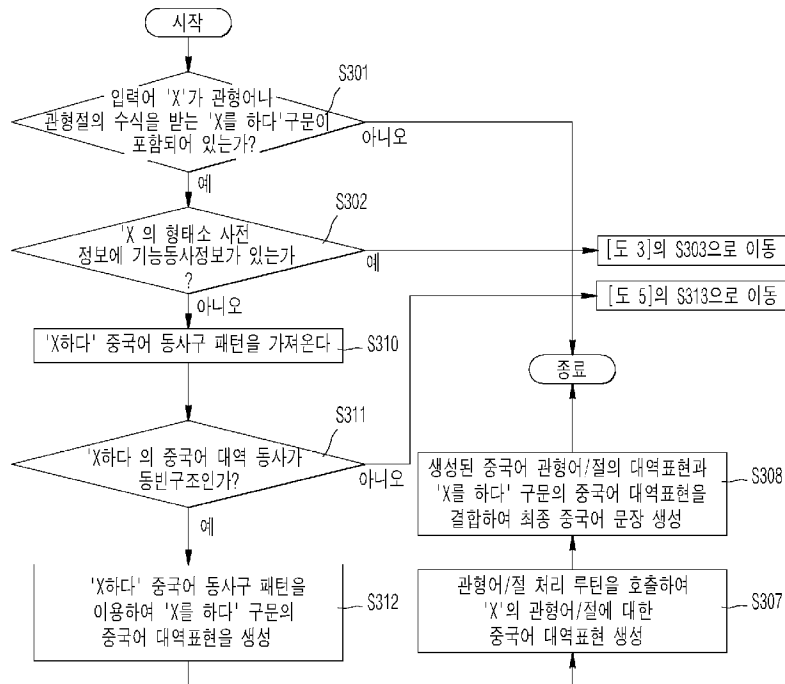
도면2



도면3



도면4



도면5

